DOCUMENT RESUME

ED 229 402 TM 830 298

AUTHOR Scheerens, J.; van Seventer, C. W.
TITLE Political and Organizational Pre-Conditions for Application of the Standards for Educational Evaluation.
PUB DATE Apr 83
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Committees; *Evaluation Criteria; *Evaluation Needs; Foreign Countries; *Organizational Climate; Policy Formation; *Political Influences; Program Development; Program Evaluation; *Standards
IDENTIFIERS *Netherlands; *Standards for Evaluation Educ Prog Proj Materials

ABSTRACT

The "Standards for Evaluations of Educational Programs Projects Materials," formulated by the Joint Committee on Standards for Educational Evaluation raise certain questions which should be regarded as indicating a further complementary perspective rather than as a criticism of their formulation. The possibility of conflict between the various standards, their scope, and the existence of political and organizational context factors in evaluation research suggests that the whole evaluation problem should be considered at a higher level of aggregation; the level of national research policy and in the institutional and organizational structures at that level. In the conceptualization of policy and evaluation research and their relationship, three propositions in the policy-making and organizational context of evaluation research lead to a whole set of problems raised by certain standards. These are "conflicting interests," "political viability," and various standards to do with the evaluation object, groups and persons involved: object identification, public identification, and information scope and identification. A brief account of a study of the functioning of policy-oriented research in some innovative programs in Dutch education elucidates why the problems should be looked at primarily in connection with management-organizational conditions at a macro-level. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

D 83-501

ED229402

Political and organizational pre-conditions for application of the Standards for Educational Evaluation.

(the 'standards for educational evaluation' seen in a wider perspective in relation to policy-making and organization).

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Scheerens

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

J. Scheerens and C.W. van Seventer
SVO (Foundation for Educational Research in the Netherlands).

Introduction

In the United States, over the period from 1975 to 1981, a group of evaluation experts led by D.L. Stufflebeam set themselves to formulate a series of quality criteria or 'standards' for evaluation research in education. The final results of this work, which was carried out on a wide scale, appeared in the report of the Joint Committee 1981 which proposed thirty standards spread over four main categories. These categories are as follows: utility standards, feasibility standards, propriety and accuracy standards. For example, under utility would come the degree of clarity of reports and their punctuality; the political viability of the research would come under feasibility; the public's right to be kept informed illustrates propriety and reliable measurement would conform to the accuracy standard. A clear and concise definition of each standard is given, followed by a further explanation and an example. At the same time concrete guidelines for application of the standards are provided, together with warnings of the pitfalls and traps which may be sprung by mistaken application.

TM 830 298

Thus political viability receives the following description: evaluation research is politically viable in so far as its objectives can be achieved in the face of the pressures which can be brought to bear by various interested groups. One of the guidelines for optimal realisation of this requirement reads: contracts should be negotiated in which mutual rights and duties are fixed and which guarantee the evaluators access to

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

the information they need. In the context of political viability a warning is given against creating the impression that the evaluation research is bound, a priori, to work out to the advantage of a particular interest group. Discussion of this standard closes with a description of a case where evaluators found that by giving way to political pressure from a certain interest group they had given certain influential representatives of an opposing group a public stick with which to beat them. (See Joint Committee, pp. 56 - 59).

Although, as the symposium on the Standards held during the 1980 AERA congress showed, a certain apprehension remains in certain quarters that the standards may act as a straitjacket precisely at a time when evaluation research is crying out for flexibility in the use of available strategies, techniques and creativity, it is our opinion that the standards have an undoubted contribution to make to the professionalisation of evaluation research. In this connection we subscribe fully to the view of Cronbach and his co-workers (Cronbach et al., 1980) when they say that professionalisation is a necessary condition for the improvement of evaluation practice. This is particularly true for the Netherlands which already suffers from the fact that its empirical tradition in the social sciences and also specifically in the educational framework, is considerably weaker than that of the United States.

A closer analysis of the standards nevertheless raises a certain number of questions which we shall later show should be regarded as indicating a further complementary perspective in which to look at the standards rather than as a criticism of their formulation.

One of the first points to be noticed is that some of the standards seem to conflict with each other. For example, in the interest of the validity and reliability of the research it may be best to keep those who are to be the sources of information in partial ignorance of some aspects of the program, for instance how the techniques (interviews, schedules and questionnaires) of research have been built up and assembled. In this way standards of accuracy lead to a breach of what is laid down by the propriety standards which is that the public has the right to be kept informed on the evaluation procedures. In view of the fact that the Joint

Committee report takes account of this conflict (p. 9) it will not be dealt with further here. The possibility of conflict between the various standards emphasises once more that these quality requirements do not lend themselves to mechanical application but should rather be weighed up in relation to all sorts of context aspects in any practical situation.

The second striking point about the standards is their scope. This does justice to the complexity of the whole problem of evaluation into which enter questions of an ethical, political, organizational, juridical and research-technical nature. This many-sidedness does however raise the question whether evaluation researchers are being asked to perform the impossible. It has already been remarked by A.D. de Groot that action-research was 'too difficult for anyone' (De Groot, 1975). The question is whether, for example, training programs should have to rely on evaluation researchers having as it were twelve pairs of hands. The report of the Joint Committee devotes little space to the most obvious solution to the problem which would be to work with officially recognised multi-disciplinary evaluation teams whose members would be specialists in various different fields.
On this point it is our opinion that the whole evaluation problem should also be tackled at a higher level of aggregation which is to say at the level of national research policy and in the institutional and organizational structures belonging to that level. This is the key issue in the remaining part of this article.

The third question is whether, given the existence of political and organizational context factors in evaluation research, the specific view of the individual evaluation project used in the 'Standards' should be supplemented and if so to what extent. For various reasons we think that policy-making and organizational aspects of the evaluation problem should be considered at a higher level of aggregation. The literature of evaluation research (particularly policy-oriented research) speaks more and more of the relation between policy-making and the practice of the research which is supposed to contribute to the formulation of the policy (see, for example, Weiss, 1975; Lindblom and Cohen, 1979; Cronbach et al., 1980).

In our opinion this literature can be said to show that it is extremely important for teams of evaluation researchers to have the opportunity to set up the research against the background of as basic an analysis as possible of the policy-making processes in question. Furthermore, the results of a recently completed study of the functioning of policy-oriented evaluation research in the context of a number of innovatory programs in Dutch education, indicate that characteristics of the national education system and patterns of organization play an important role in the harmonisation of evaluation research and policy (Scheerens, 1983[a]). We consider this method of approach to the evaluation problem as complementary to the attitude dominant in the 'Standards'. The following section throws further light on this. The extra element in our point of view is that it concerns itself with a higher level of aggregation; some of the quality requirements, as formulated by the Joint Committee, can simply be 'translated' to the new level without much difficulty, as will be seen later on.

On the policy-making and organizational context of evaluation research

Since the second half of the seventies the general body of thought on policy and program evaluation has produced two central schemes for discussion: the rational decision model (for the characterising of policy-making) and the experimental design (as prototype for setting up evaluation research). There has also been a corresponding change in the notion of the relationship of policy-making to evaluation research. The 'classic' concept of 'reforms as experiments' (compare Campbell, 1969), assumes a direct relationship between evaluation research and policy. According to this view the results of the research are considered as giving direction to the process of taking policy decisions. This concept of the connection between policy-making and evaluation research is distinguished as the instrumental or linear model (compare for example Weiss, 1982). Against this is set the idea of a less rigid linkage between policy and evaluation research according to which policy-making is seen as a diffuse process of small 'adjustments' and evaluation research has, at the most, an indirect, partial and long-term influence as the ideas and concepts held by policy-makers gradually adjust to it. This way of looking at the relationship is known as the 'enlightenment model'. The 'rational policy-making model' as prototype for policy-making is replaced in this scheme by Lindblom's concept of incrementalism. At the same time, in evaluation literature of a more methodological and research-technical kind there has been a 'qualitative' trend which protests just as vigorously against experimental research design as the authors who see policy-making as incrementalism kick against the rational decision model (see for example Parlett and Hamilton, 1972).

It would be beyond the scope of this article to embark on a detailed analysis of the conceptualisation of policy and evaluation research and the relation between them. We have chosen instead to put forward our conception of the subject as three propositions (for more detailed argumentation see Scheerens 1983[b]).

a. Neither the typifying of policy nor the choice of an evaluation design must be seen as a choice between two extremes. In weighing up evaluation strategies as well as in the interpretation of policy-making according to formal models, there are a number of intermediate 'stations'. In the case of policy-making may be mentioned Etzioni's concept of mixed scanning and the various ways of completing Simon's model of 'bounded rationality'. Policy-making models may be sub-divided according to the information they require about means-ends relations and complexity and scale, in the sense of extent and length in time (compare Scheerens, 1983[b]).
Similarly, the choice of an evaluation design does not have to be limited to experimental versus informal, qualitative and descriptive. Various quasi-experimental designs, non-experimental causal analysis and mutually comparable case-studies, for example, lie between the two.

The characterisation of the policy-making process to which the evaluation research should contribute, presents salient points which can be seized upon in order to arrive at the choice of a specific evaluation design. Policy-making models thought of as arranged on a continuum from 'rational' to 'incremental' can be matched with evaluation designs along a continuum from experimental to descriptive ex post facto research (compare Scheerens, 1983[b], ch.3). An important organizational pre-condition for reaching the best possible dovetailing of policy-making to the evaluation research structure is that the adjustment process should be controlled (for example, where both policy-makers and evaluation researchers recognise the problem the process might then be executed by a special mediating body).

b. The typifying of policy-making processes according to certain decisional models has the character of a theoretical construction. A policy that is at first of a diffuse character acquires a more distinct form when its contours are more closely plotted and its priorities, objectives and means more precisely formulated. Researchers in the field of evaluation who are attempting to reach a greater degree of certainty about the nature of policy-making processes, within which framework evaluation research has to take place, are enabled, in this way, to have a part in influencing policy

in its formal aspects. In connection with this we would like to point to the great heuristical significance of the rational decision model. In the case of fragmentary and diffuse policy-making as well it will be important to trace means-ends relations be it only in a rough and ready way. The recognition of such a possibility of the active shaping of diffuse policy-making processes by evaluation researchers, brings with it important consequences for management and organization. On one hand evaluation teams must have sufficient independence to be able, if necessary, to tackle construction or reconstruction but on the other hand care must be taken that these constructions are approved (which means that they must, for instance, be laid before different groups of people concerned in the program).

c. Organizational pre-conditions are also of crucial significance in a more general sense, because they are essential to an improved coupling of policy-making with policy-oriented research. Here, power relationships are of central importance. Some authors even see the present preference for qualitative, open and formative-oriented evaluation approaches as the 'ideological superstructure' thrown up by researchers in a 'conspiracy' with those who carry out the programs that are to be evaluated (compare Hofstee, 1982), a conspiracy which is the result of the constant marked dependence of researchers on those who actually carry out the program.

The preceding considerations of the policy-making and organizational context of evaluation research lead in the first place to an underlining of the importance of the whole set of problems raised by a certain number of the standards. The particular standards in question are 'conflicting interests', 'political viability' and various standards to do with the identification of the evaluation object and the most important groups and persons involed ('object identification', 'public identification' and information scope and identification;'). Using a brief account of the principal results of a recently completed study of the functioning of policy-oriented research in the context of some innovatory programs in Dutch education, the following section will attempt to elucidate why the set of problems we have mentioned should be looked at primarily in connection with 'macro aspects': characteristics of the management-organizational context at national level.

The importance of organizational conditions at macro level; an example

The object of the study mentioned above was to form a 'retrospective assessment' of what had been experienced in policy-oriented (evaluation) research as carried out since 1976 in the context of some national innovatory programs. Three of the four 'large' innovatory programs were analysed in the study: the Middle School, Open School and Integrated Primary School experiments (the innovatory program for participatory education - part-time education for young adults - was disregarded). The study attempts to describe and analyse the research in its policy-making and organizational context. It is not only an appreciation of research programs but at the same time gives an account of the policy-making in each of the sectors and of the organizational configurations within which the research was planned and carried out. The data were gathered by means of document analysis and interviews with key people (departemental policy planners, experts on innovatory programs and teachers). A number of research reports were examined by external experts; furthermore, some of the written accounts of draft research programs were submitted to the researchers concerned.

The results of the study painted a rather less than rose-coloured picture of how the research in question was functioning. The key people who were interviewed gave an overwhelmingly negative assessment of the relevance to policy of most of the research programs. The external experts consulted were extremely critical of the significance of a selected number of programs and the quality of the research technique.

The way the research programs ran showed that there was a lot wrong with planning and research management, among other things the length of the procedures in preparation for the research and the difficulties of co-operation between schools and researchers. In a number of cases these two groups came into conflict over the choice of instruments and the way in which these could be administered. This led to a loss of respondents, the relinquishing of the use of instruments and often to the administration of instruments under less than strictly controlled conditions. The context of the research, from the point of view of policy and organization, was different in each of the three innovatory programs. In the Integrated Primary School program conditions were on the whole more favourable than in the other two programs. It was also in this sector

that the character of the research done was such as to indicate that it would be an exaggeration to attribute the preponderance of disappointing results solely to the influence of external conditions. Particularly where research-management of the separate programs is concerned, there is good reason to take the results of this study as an encouragement to embark on greater professionalisation of evaluation researchers.
Yet in our interpretation of the data we feel it necessary to lay the greatest accent on the influence of external circumstances on the policy-making and organizational content of the research concerned.
To put it bluntly, even if each separate program team had acted in strict conformity with the guidelines provided by the 'standards of educational evaluation', the problems experienced would still only have been reduced by about half. More precisely, the possibilities of living up to the standards at program level are severely limited by factors in the larger context of organization and management. The degree to which an evaluation team itself contributes to guaranteeing a program's 'political viability' is partly dependent on its own position within the network of official bodies involved.

In the practical situation which we are considering we are of the opinion that the context factors of real significance for the way in which the research in question was conceived and executed are as follows:

1. The innovatory programs in education with which we are dealing, are to be seen as contrivances for pursuing a policy of educational reform in the Netherlands shaped more than ever before by government initiatives and the influence of external education experts. At the same time the general spirit of opinion over the same period was largely permeated by the ideology of democratisation from the bottom upwards. The partly petrified and 'segregated' Dutch education system itself, the legally-recognised and relatively important autonomy of the individual schools (and within the schools the professional autonomy of the teaching staff) are all pre-emptive factors for any projected education policy (Idenburg's conception of the constructive politics of education and see also Leune, 1981).

2. Although the innovatory programs were given the name of 'experiments', there was a great deal of doubt as to the exact signification of this appellation. The Second Chamber debated the question of whether the Middle School experiment was supposed to show if Middle Schools were to be introduced or merely to establish how the idea of Middle Schools might be realised in practice. The conclusion of this debate was that it was indeed a question of whether Middle Schools should be brought into existence. However, the principal architect of the Middle School experiment, the Innovatory Commission on Middle Schools, has never accepted this conclusion. This difference of opinions on the status of the innovatory education programs was naturally of great significance for the place allotted to evaluation.
3. In spite of the national character of the innovatory programs, the use of education experts further to elaborate the policy's points of departure, and the projected management role of the central authority, the innovatory strategy chosen was one which left the initiative largely in the hands of the individual schools involved in the experiment. The widely divergent developments which were the consequence created problems for a policy-oriented research that was supposed to provide data which could be combined to cover all the separate schools.
4. It appears from the literature on organization that professionals such as teachers have, on the whole, a tendency to resist any evaluation of their work. Within the Dutch education system schools and teaching staff enjoy considerable autonomy. An innovatory strategy which puts the accent on the basis (which is to say the teaching level) may well tend to increase the teachers' autonomy even more. All these factors put together mean that in negotiations between teachers and researchers on the execution of the research, the researchers often get a poor deal and have to be satisfied with research procedures which they themselves consider far from ideal.
5. The structure of the organization which surrounds research in the context of innovatory programs in education has the features of an interorganizational network in which the relationships of the components are not completely formalised. On this level as well strings are constantly being pulled in attempts at manipulation and power-seizing in research. To illustrate this point we can refer to

the differences of opinion between, for example, the Foundation for Educational Research, which, in the Netherlands is the central organ responsible for managing a large proportion of Dutch educational research, and the Minister of Education. On several occasions proposals for research programs that the Foundation judged as below par were nevertheless approved and set in train by the Minister. In such a complex organizational configuration institutionalised research in education occupies a comparatively weak position.

In our opinion the factors already mentioned complete the background for the claim that such research may justifiably be seen as a plaything bandied to and fro between the various interest groups. Our view is that first of all a national education policy on the professionalisation of evaluation research must be instituted in order to create the pre-conditions in which a professionalisation more real in itself can be brought to flourish. In the closing sections we shall describe some measures by which this might be achieved.

Conclusion : pre-conditions of management and organisation in the application of the 'standards'.

Earlier on we came to the conclusion that certain 'standards' were of special significance when looked at in the light of recent contributions to the discussion of evaluation research (conflicting interests, political viability and various 'indentification' standards). The preceding section gave an example to show why we consider management-organizational conditions at a 'macro-level' particularly significant for the improvement of evaluation practice. Any national research policy aimed at professionalisation in the spirit of the standards will at the same time have to pay attention to improvement of these conditions. The possible measures for creating such a research policy, which we shall now describe, will serve to illustrate this idea. As a basic regulating principal we can take the 'Standards' as formulated by the Joint Committee, at first enumerating a number of conditions for professionalism and then 'translating' some of the standards into policy measures at a national level. The feasibility of this last excercise demonstrates that at least some of the standards can indeed be considered as quality requirements for a national policy on policy-oriented evaluation research.

measures which stimulate application of the standards

In the first place training and further training of evaluation researchers should also be taken note of in a national education policy; for example by providing grants, the organization of international workshops and courses.
In the second place the importance of assessing completed evaluation research must be stressed.
In the third place what must be fought for is a strengthening of researchers' work communities and of professional organizations. In such a small country as the Netherlands the opportunity of keeping in contact with foreign evaluation organizations is extremely important to researchers. Finally, the attempt must be made to see that other groups of people involved, such as policy-planners and those who carry out the programs to be evaluated are also included in certain additional training courses.

conflicting interests and political feasibility

What emerges once more from the study of which we have given a brief résumé in this article is the significance of conflicting interests for evaluation research and their threat to its political viability, in the sense given by the 'Standards'. In general we regard the taking of measures to strengthen the independence of national coordinating institutions, educational research foundations and research teams in the field as of top priority in order to combat the many problems of these context aspects. It is clear that this is no mean task, particularly at a time when budgets are being whittled down as far as possible. In addition, a certain defense might be erected against political bias and 'misuse', of research, by formalisation of the relations between schools, researchers and research funders and clients; this might be looked on as an application of 'formal obligations' - one of the 'standards' - but on a national scale.

Finally, evaluation models which seek to express conflicting interests about the object of the evaluation research are of great importance in this connection. An example is the 'betting model' developed by Hofstee (Hofstee, 1982).

various 'identification' factors

The literature in which the diffuse character of policy-making processes is emphasised also establishes the importance of 'identification processes'. We have already spoken of the evaluation researcher's task of constructing or reconstructing policy-making processes. Organizational measures which promote the identification of the characteristics of subject and context of the evaluation research, are the planning of preparatory activities in order to arrive at choices which are as solidly grounded as possible, and the possible setting-up of intermediate organs such as programming committees.

Also, in connection with this aspect, some degree of formalisation of the relationships between policy planners, researchers and those who execute the programs which are to be evaluated, is to be recommended, for instance the drawing up of agreements on co-operation in the most explicit possible form.

Further professionalisation of those involved in evaluation research and the creation of conditions of organization and management which will stimulate this process and improve the practice of evaluation research are to be seen as mutually reinforcing processes. In particular, we would urge that an evaluation community which was more professional would itself develop a stronger 'political will' in order to insist on improvement of conditions for research.

BIBLIOGRAPHY

Campbell, D.T., Reforms as Experiments. American Psychologist, vol. 24, no. 4, 1969.

Cronbach, L.J. and Associates, Toward Reform of Program Evaluation, Aims, Methods and Institutional Arrangements. San Francisco: Jossey Bass Publishers, 1980.

Groot, A.D. de, Bijdrage aan symposium over aktie-onderzoek, onderwijsresearchdagen, Utrecht, 1975.

Hofstee, W.K.B.. Evaluatie: een Methodologische Analyse. Tijdschrift voor Onderwijsresearch. Jrg. 7, no. 5, september 1982.

Joint Committee on Standards for Educational Evaluation, Standards for Evaluations of Educational Programs and Materials. New York: Holt, Rinehart & Winston, 1981.

Leune, J.M.G., Besluitvorming in het Onderwijsbestel. In: J.A. van Kemenade (red.) Onderwijs: Bestel en Beleid. Groningen: Wolters-Noordhoff, 1981.

Lindblom, C.E. and Cohen, D.K. Usable Knowledge. Social Science and Social Problem Solving. New Haven and London: Yale University Press, 1979.

Parlett, M. and D. Hamilton, Evaluation as illumination: a new approach to the study of innovatory programs. Occasional Paper, University of Edinburgh, Oct. 1972.

Scheerens, J., Evaluatie-onderzoek en Beleid. Methodologische en Organisatorische Aspecten. Harlingen: Flevodruk, 1983[a] (no. 68 SVO-reeks).

Scheerens J. (red.), Het sector-onderzoek: onderwijsonderzoek in de marge van beleid en wetenschap? Een studie naar het functioneren van het beleidsgerichte onderzoek in het kader van enkele landelijke onderwijs-innovatieprojecten. Harlingen: Flevodruk, 1983[b] (no. 69 SVO-reeks).

Weiss, C.H., Improving the Linkage between Social Research and Public Policy. In: Lynn, L.E. (ed.) Knowledge and Policy: the uncertain Connection. Study project on Social Research and Development, vol. 5, 1975.

Weiss, C.H., Policy Research in the Context of Diffuse Decision-Making. In: Kallen, D.B.P. and others (eds.). Social Science Research and Public Policy-Making: A Reappraisal, NFER-Nelson Windsor, U.K., 1982.